Three-dimensional thyroid assessment from untracked 2D ultrasound clips

Wolfgang Wein[1], Mattia Lupetti[1], Oliver Zettinig[1], Simon Jagoda[1], Mehrdad Salehi[1], Viktoria Markova[1], Dornoosh Zonoobi[2], and Raphael Prevost[1]

[1] ImFusion GmbH, Munich, Germany
[2] MEDO.ai, Alberta, Canada

Abstract. The diagnostic quantification of thyroid gland, mostly based on its volume, is commonly done by ultrasound. Typically, three orthogonal length measurements on 2D images are used to estimate the thyroid volume from an ellipsoid approximation, which may vary substantially from its true shape. In this work, we propose a more accurate direct volume determination using 3D reconstructions from two freehand clips in transverse and sagittal directions. A deep learning based trajectory estimation on individual clips is followed by an image-based 3D model optimization of the overlapping transverse and sagittal image data. The image data and automatic thyroid segmentation are then reconstructed and compared in 3D space. The algorithm is tested on 200 pairs of sweeps, and shows that it can provide fully automated, but also more accurate and consistent volume estimations than the standard ellipsoid method, with a median volume error of 11%.

1 Introduction

Ultrasound imaging has been for a long time the gold standard for thyroid assessment, thus replacing clinical inspection and palpation [1]. The current workflow is to perform 2D measurements on ultrasound planes and combine them into a volume using the so-called ellipsoid formula, a very coarse approximation which leads to uncertainties [11] and sub-optimal reproducibility. In this context, creating a 3D reconstruction of the thyroid would bring several benefits: (i) it would allow a more precise volume estimation of either thyroid or suspicious masses than 2D imaging [6], and (ii) it could ease matching of the same anatomy, enabling easier and more reliable regular screening of the population at risk. One way to achieve this could be 3D ultrasound solutions such as a dedicated 3D transducer, or an external tracking of a 2D probe [13]. Such approaches have not found widespread adoption though, mostly due to the high cost or cumbersome setup.

It has recently been shown that it is possible to reconstruct the 3D trajectory of a freehand ultrasound clip using deep learning [8]. Combined with a thyroid segmentation method (see for instance [2, 5] for deep learning-based approaches,

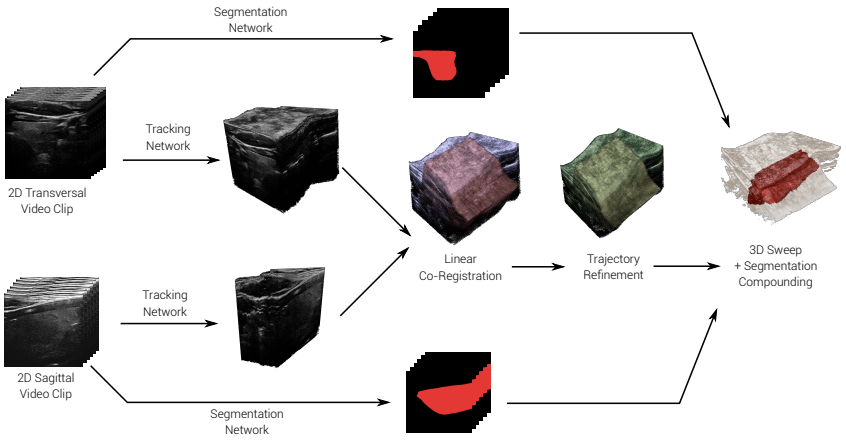

Fig. 1. Overview of the proposed method for 3D thyroid reconstruction from two perpendicular sweeps.

or [3] for a more generic and thorough review), this could enable a better assessment of its volume. However, this approach might still suffer from drift degrading the accuracy of such a measurement. In this paper, we therefore propose to build upon such methods by exploiting the redundancy from two perpendicular freehand acquisitions: out-of-plane distances in one sweep can be precisely recovered from the other sweep since they appear in-plane. In order to do so, our approach will aim at registering those two acquisitions while jointly refining their own trajectories. This is to our knowledge the first time a consistent 3D reconstruction of the thyroid on untracked freehand 2D ultrasound clips is presented, accurate enough to utilize volumetric lesion measurements.

2 Methods

Our overall approach takes two ultrasound clips as input, slow-swept in transversal (TRX) and sagittal (SAG) directions over one side of the thyroid gland (right or left). The output are two registered volumetric representations which are visually presented to the clinician, along with a 3D thyroid gland segmentation. An overview of the computational pipeline is shown in Figure 1, each step of which is described below.

2.1 Deep Learning Tracking Estimation

Similarly to [9], we first train a convolutional neural network (CNN) to estimate the trajectory of the probe during the sweep based on the video clip. The network uses as input a 4-channel image (two successive ultrasound frames, and the

2D optical flow in-plane motion between them encoded as two channels) and produces the 6 pose parameters (3 for translation, 3 for rotation) of the relative 3D motion. No changes have been introduced to the network architecture or learning parameters compared to [9]. When applied for all successive frame pairs and therefore when accumulated, the entire trajectory can be reconstructed. Even though this method is able to capture the global motion of the probe, it might still yield some drift which would significantly degrade the estimation of the volume. We therefore rely on the complementary information from two perpendicular sweeps to fix this potential drift.

2.2 Joint Co-Registration & Reconstruction

Initial Registration With both the transversal and sagittal sweep available in a first 3D representation, a set of 3D volumes in a Cartesian grid is created of both the B-Mode intensities and their pixel-wise labeling, using an efficient GPU-based compounding algorithm similar to [4]. An initial rigid transformation between them is derived by assuming standardized scan directions, as well as by matching the segmented thyroid volumes with a rigid registration. This rigid transformation is further optimized by maximizing a cross-correlation metric over the image intensities. This aligns the bulk of the anatomical structures roughly, despite possibly incomplete thyroid visibility (and hence segmentation) in both sweeps. Most of the drift by the initial sensorless reconstruction method cannot really be fixed at this point, since the out-of-plane lengths of the sweeps stay fixed.

Trajectory Refinement via Co-Registration The reconstruction of both sweeps is then simultaneously refined by optimizing a number of trajectory parameters, together with the rigid transformation between the sweep centers, with respect to an image similarity on the B-Mode intensities. Since the 3D geometry within an ultrasound volume changes during this optimization, an on-the-fly multi-planar reconstruction (MPR) method is required to compare individual frames from one sweep with a compounded image from the other one. A related technique was presented in [12] for image-based optimization of the probe-to-sensor calibration for externally tracked 3D freehand ultrasound. We generalize it to optimize any parameters affecting the 3D topology of un-tracked data. A cascade of non-linear optimization using a Nelder-Mead search method are executed with increasing degrees of freedom (DOF) parameters as follows:

- **10 DOF:** Relative rigid pose plus 2 DOF per sweep, corresponding to an additional out-of-plane stretch, and rotation around the probe surface between the first and last frames of the sweep, as drift correction from the sensorless reconstruction method.
- **18 DOF:** As above, with the 2 DOF per sweep replaced by a 6-DOF rigid pose between the first and last frame.
- **30 DOF:** As above, with an additional 6 DOF pose control point in the center of the sweeps, realized through Hermite cubic spline interpolation on the individual rigid pose parameters.

– **54 DOF:** As above, with 3 control points per sweep instead of one, placed in an equidistant manner over the number of image frames.

2.3 Automated Thyroid Segmentation

Thyroid gland segmentation has been studied for many years in the literature. Following the recommendation in the review [3], we opted to segment each 2D ultrasound frame using a very standard U-Net [10] neural network. Using the same trajectory found in the previous steps, we then compound those label maps in both transversal and sagittal sweeps into 3D binary masks. After minor post-processing (morphological closing), the final segmentation is defined as the union of the two 3D segmentations, coming respectively from the transversal and corresponding sagittal sweep.

The whole pipeline is implemented in the C++ ImFusion SDK, with OpenGL shaders for similarity measure computation and image compounding, and CUDA for the deep learning models. The overall computation time on a standard desktop computer is in average 3 minutes, the deep learning based segmentation and tracking estimation take approximately 10 seconds per sweep, while the successive stages of the joint sweep reconstruction take from 20 seconds to 1 minute. The computation time can be further reduced by running the tracking estimation and the segmentation in the background during the acquisition.

3 Experiments and Results

3.1 Data Acquisition

Our method is evaluated using a dataset of 180 ultrasound sweeps from 9 volunteers acquired from a Cicada research ultrasound machine (Cephasonics, Inc., Santa Clara, CA, USA) with a linear probe. For each volunteer, we acquired 5 transversal and 5 sagittal sweeps for each lobe of the thyroid. The aquisitions were performed by three different operators (the transversal top to bottom and the sagittal starting from the trachea) with variations in the aquisition speed, extend of the captured anatomy, tilt angles. Following guidelines in [8], the sweeps have been tracked using an optical tracking system and recorded without speckle reduction or scanline conversion.

In order to train and evaluate our method, the thyroid was manually segmented by several operators on a subset of the sweeps. The volunteers were then split into a training (5) and testing (4) subset; the first set was used to train all networks as well as fine-tune the registration parameters and contains 100 sweeps, the latter was left out for the evaluation of the method and contains 80 sweeps.

3.2 Experiments

Our evaluation design is driven by the 3D thyroid segmentation, in particular its consistency between TRX and SAG sweeps, as well as the volume estimation

since this is the relevant clinical measure. We ran our whole pipeline with different configurations to test our various hypotheses. All results are summarized in Table 1 and discussed below.

Table 1. Results of the experiments on co-registered sweeps without tracking information. The Dice coefficient TRX/SAG represents the overlap between the 3D segmentation of the thyroidal gland computed from the transversal and sagittal sweeps. This number is used as quality metric for all the sweep co-registration methods presented. As reference the Dice overlap computed with the ground truth tracking is reported. All numbers are means ± standard deviation if not otherwise specified.

exp. #	sweep co-registration	Dice TRX/SAG mean ± std (median)	volume error	norm. vol. error	rel. trajectory error	rel. length error
1	ground truth tracking	0.69 ± 0.13 (0.72)	N/A	N/A	N/A	N/A
2	rigid reg.	0.61 ± 0.14 (0.63)	2.09 ± 1.44 mL	0.25 ± 0.13	0.16 ± 0.06	0.22 ± 0.13
3	10-DOF refinement	0.64 ± 0.12 (0.65)	1.22 ± 1.46 mL	0.14 ± 0.16	0.16 ± 0.06	0.17 ± 0.14
4	18-DOF refinement	0.64 ± 0.12 (0.66)	1.14 ± 1.48 mL	0.14 ± 0.16	0.17 ± 0.09	0.16 ± 0.14
5	30-DOF refinement	0.65 ± 0.12 (0.66)	1.15 ± 1.47 mL	0.14 ± 0.16	0.16 ± 0.09	0.16 ± 0.14
6	54-DOF refinement	0.66 ± 0.12 (0.68)	1.15 ± 1.45 mL	0.14 ± 0.16	0.16 ± 0.09	0.16 ± 0.14

Thyroid Segmentation The ground truth thyroid segmentations were annotated by a single non-expert operator on 58 sweeps, 50 of them from the 5 volunteer training subset and the remaining 8 from the 4 volunteer testing subset (1 pair per lobe per test volunteer). The segmentation U-Net was trained and evaluated on the first 50 sweeps from training subset. The segmentation U-Net achieves a median Dice coefficient of 0.85 on our validation set of 2D frames.

After 3D compounding of the individual frames, the Dice coefficient between manual segmentation and network output drops to an average of 0.73 ± 0.08 because of inconsistencies across slices or ambiguity near the ends of the thyroid, which is in agreement with the scores recently reported in [13]. Due to the fact that each orthogonal sweep captures a slightly different view of the thyroid, also the Dice coefficient between compounded TRX and compounded SAG sweeps with manual annotation is not 1, but rather around 0.70 ± 0.05. Since those two numbers are in the same range, we then assume that, in the context of registration evaluation and because segmentation is not the focus of the paper, metrics on the network output are a good proxy for metrics on manual segmentations. This allows us to consider all the 200 possible pairs of test sweeps in the next experiments (25 pairs per lobe per test volunteer) instead of the 8 which are manually labeled (1 pair per lobe per test volunteer).

Single Sweep vs Multi-Sweep Due to the residual drift of the tracking estimation network, evaluating the thyroid volume from a single untracked sweep produces inaccurate estimates with, according to our early experiments, a 45% error. We therefore conclude that a second perpendicular sweep is required to bring the missing out-of-plane information. In the experiments #2 to #6, we use

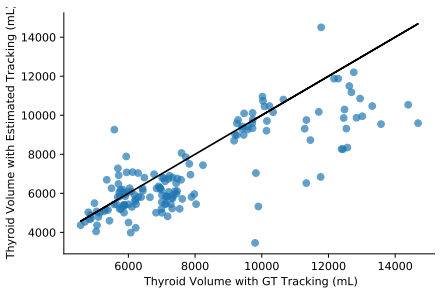

Fig. 2. Comparison of the estimated volume with ground truth tracking versus estimated trajectory on the 200 pairs of sweeps (Spearman correlation = 0.75).

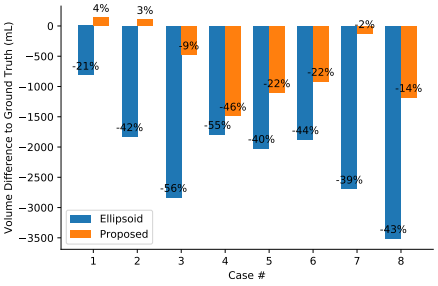

Fig. 3. Volume difference between true thyroid volume and estimated volume for both the ellipsoid method and our proposed approach on the 8 cases with manual segmentation.

two such sweeps, which indeed yields a significant improvement of all reported metrics.

Trajectory Refinement We evaluated the thyroid volume assessment at different stages of our refinement pipeline (from 10 to 54 DOF). Our experiments #3 to #6 demonstrate the benefit of further adjusting the trajectory of the two sweeps using our joint registration/trajectory refinement approach. According to a Wilcoxon statistical-test with a p-value threshold of 0.01, differences become irrelevant for motion corrections with more than 18 DOF. Furthermore, we compared the estimated sweep trajectories at each stage with their corresponding ground truth trajectories. To this purpose we defined the relative trajectory error as the cumulative in-plane translation error at each sweep frame divided by the ground truth sweep length, and similarly the relative length error as the relative error of the estimated sweep length w.r.t the ground truth length. The values of the relative trajectory and length errors are reported in the two rightmost columns of Table 1. With the same statistical-test settings as in the thyroid volume assessment, we conclude that the joint registration significantly improves the initially estimated overall length drift of the sweep, while there is no substantial change in the local trajectory estimation. The latter can be attributed to some extent to the error of the ground truth external tracking, i.e. the true local probe & tissue motion is not known with sufficient accuracy.

Figure 2 shows a comparison of the thyroid volume between the ground truth tracking and the estimated tracking. The two quantities are in strong agreement, with a Spearman correlation of 0.75, although the estimated volumes tends to be smaller than the ground truth ones. This is due to the fact that the tracking estimation network tends to underestimate the out-of-plane motion between frames on unseen sweeps, an issue that we will further investigate in future works.

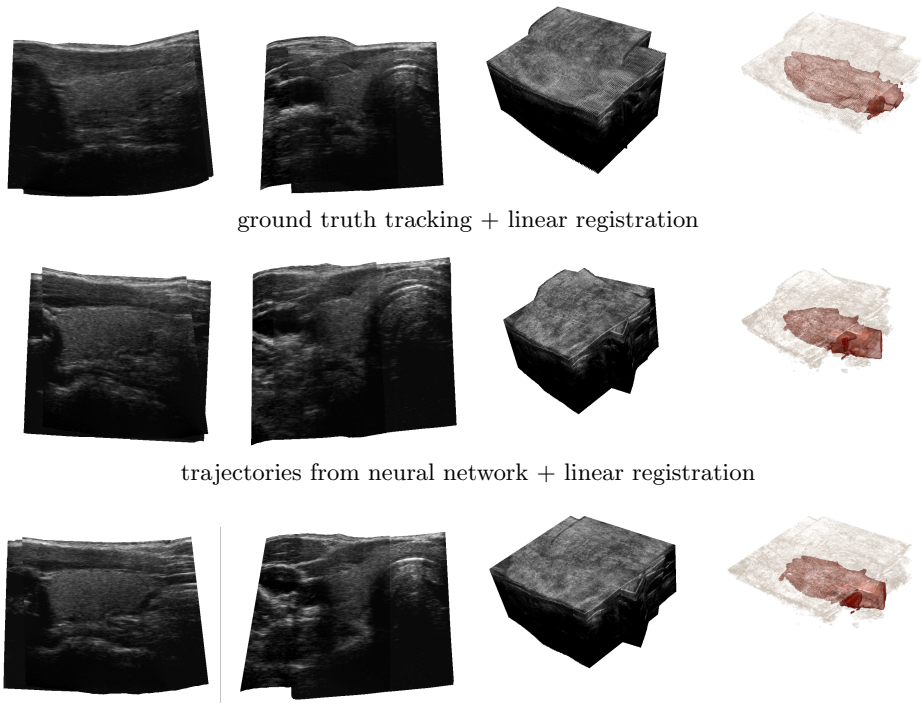

Fig. 4. 3D Reconstruction of two perpendicular sweeps of a thyroid. The four columns respectively show: (i)/(ii) blending of the two sweeps in two perpendicular planes, (iii) the registration of the two sweeps, (iv) 3D view of the resulting segmentation.

To further demonstrate the reliability of our volume estimation approach, we compare it against the current clinical workflow, i.e. the ellipsoid formula, which consists in measuring the three dimensions of the thyroid in two perpendicular planes and multiplying them with a factor of 0.529 (see [11]). We report in Figure 3 our final volume estimation as well as the ellipsoid-based estimate, compared to the ground truth segmentation, on the 8 cases for which it is available. Our method yields more consistent and accurate estimates than the current clinical standard (median error of 12% vs 42%).

Finally, we show in Figure 4 qualitative results on a representative case of our dataset, where we notice a stronger agreement between the two sweeps after our trajectory refinement. Here, the visual appearance is arguably even better than for the ground truth because internal tissue motion can be partially compensated, which is not possible with external tracking alone. This also illustrates that there is an inherent limit in ground truth accuracy for our experimental setup.

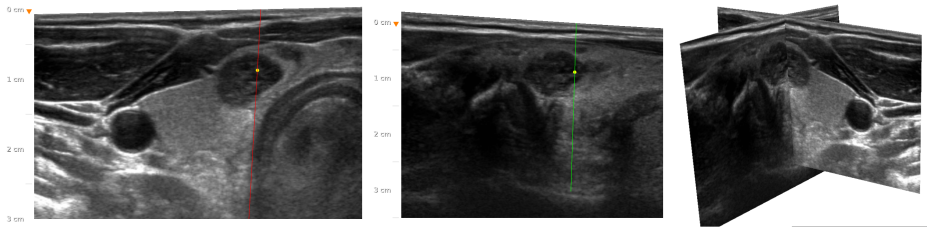

Fig. 5. Reconstructed frames from transversal and sagittal clips (left, middle), together with their 3D rendering (right) of a patient scan.

3.3 Application on Clinical Data

In an ongoing prospective study, patients undergoing screening for suspicious thyroid masses are scanned with a Philips iU22 ultrasound machine. Written consent was obtained from all patients prior to the examination. For each patient the following scans are performed:

1. Conventional thyroid ultrasound by a routine clinical protocol which includes transversal and sagittal greyscale still images and cine sweep images through both thyroid lobes with a L12-5 probe.
2. Multiple 3D ultrasound volume data acquisitions with a VL13-5 transducer through the thyroid lobe, including transversal and sagittal orientations of the probe.

Three independent radiologists are measuring the thyroid nodules in three planes with the ellipsoid formula (two faculty radiologists and a fellow radiologist, all of whom had performed over 1000 thyroid ultrasound scans prior to this study). The first cases show the vast variability in the volumetric measurements of the same nodules by different readers, thus highlighting the inaccuracy of the manual method. Figure 5 shows the result of our reconstruction pipeline on one patient; instead of guessing the relationship between axial and sagittal scan planes, we provide a linked representation and a 3D rendering, using the 3D pose of the ultrasound clip frames from both sweeps as derived by our method. This alone may improve the ellipsoid method by choosing better frames to draw consistent length measurements from; in addition, an automatic segmentation method can be directly translated to 3D volume measurements. On this patient, the shown lesion was measured by three readers to be 0.85 mL, 1.06 mL, 0.76 mL, respectively. The 3D segmentation results based on our reconstruction were 0.99 mL and 0.91 mL for transversal and sagittal, respectively.

4 Conclusion

We have presented a pipeline to create accurate three-dimensional representations of the human thyroid from overlapping 2D ultrasound clips acquired in

transversal and sagittal directions. Deep learning-based reconstruction and segmentation is performed individually on each sweep, then their information is combined and redundancy in the overlapping data exploited to refine the 3D reconstruction. Since neither specialized hardware or setup is required, this can have beneficial implications for many clinical applications; we therefore propose this concept as a general means to create a 3D representation from arbitrary DICOM clips, even when using inexpensive point-of-care ultrasound probes.

While our results are preliminary due to the size of our validation set, we believe they are sufficient proof to show that the presented approach is viable in general. Pending completion of our ongoing patient study with suspicious thyroid masses, it is straightforward to build a clinical software tool, which shall also allow Deep Learning-based segmentation of nodules in addition to the thyroid gland, as well as automated co-registration of repeated screening acquisitions, based on the methods we have presented here. Further work would be necessary to address forth- and back motion during the freehand acquisitions, using for instance an auto-correlation approach (as used for gating in [7]) which can remove duplicate content. On challenging clinical data, landmarks placed interactively on anatomical structures of interest may be used to constrain the joint co-registration & reconstruction, while at the same time increasing local accuracy. Last but not least, we are also working on a non-linear deformation model that is matching the skin surface of the two sweeps so that the varying pressure exerted onto the patient's neck during the scanning can be properly compensated.

References

1. Brown, M.C., Spencer, R.: Thyroid gland volume estimated by use of ultrasound in addition to scintigraphy. Acta radiologica: oncology, radiation, physics, biology **17**(4), 337–341 (1978)
2. Chang, C.Y., Lei, Y.F., Tseng, C.H., Shih, S.R.: Thyroid segmentation and volume estimation in ultrasound images. IEEE transactions on biomedical engineering **57**(6), 1348–1357 (2010)
3. Chen, J., You, H., Li, K.: A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images. Computer Methods and Programs in Biomedicine **185**, 105329 (2020)
4. Karamalis, A., Wein, W., Kutter, O., Navab, N.: Fast hybrid freehand ultrasound volume reconstruction. In: Miga, M., Wong, I., Kenneth, H. (eds.) Proc. of the SPIE. vol. 7261, pp. 726114–726118 (2009)
5. Kumar, V., Webb, J., Gregory, A., Meixner, D.D., Knudsen, J.M., Callstrom, M., Fatemi, M., Alizad, A.: Automated segmentation of thyroid nodule, gland, and cystic components from ultrasound images using deep learning. IEEE Access **8**, 63482–63496 (2020)
6. Lyshchik, A., Drozd, V., Reiners, C.: Accuracy of three-dimensional ultrasound for thyroid volume measurement in children and adolescents. Thyroid **14**(2), 113–120 (2004). https://doi.org/10.1089/105072504322880346, pMID: 15068625
7. O'Malley, S.M., Granada, J.F., Carlier, S., Naghavi, M., Kakadiaris, I.A.: Image-based gating of intravascular ultrasound pullback sequences. IEEE Transactions on Information Technology in Biomedicine **12**(3), 299–306 (2008)

8. Prevost, R., Salehi, M., Jagoda, S., Kumar, N., Sprung, J., Ladikos, A., Bauer, R., Zettinig, O., Wein, W.: 3d freehand ultrasound without external tracking using deep learning. Medical Image Analysis **48**, 187 – 202 (2018)
9. Prevost, R., Salehi, M., Sprung, J., Ladikos, A., Bauer, R., Wein, W.: Deep learning for sensorless 3D freehand ultrasound imaging. In: MICCAI 2017 Proceedings. pp. 628–636. Springer (Sep 2017)
10. Ronneberger, O., P.Fischer, Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). LNCS, vol. 9351, pp. 234–241. Springer (2015), (available on arXiv:1505.04597 [cs.CV])
11. Shabana, W., Peeters, E., De Maeseneer, M.: Measuring thyroid gland volume: should we change the correction factor? American Journal of Roentgenology **186**(1), 234–236 (2006)
12. Wein, W., Khamene, A.: Image-based method for in-vivo freehand ultrasound calibration. In: SPIE Medical Imaging 2008, San Diego (Feb 2008)
13. Wunderling, T., Golla, B., Poudel, P., Arens, C., Friebe, M., Hansen, C.: Comparison of thyroid segmentation techniques for 3D ultrasound. In: Styner, M.A., Angelini, E.D. (eds.) Medical Imaging 2017: Image Processing. vol. 10133, pp. 346 – 352. International Society for Optics and Photonics, SPIE (2017). https://doi.org/10.1117/12.2254234